

Progress Report on  
  
Advanced Detection and Classification  
Algorithms for Acoustic-Color-Based  
Sonar Systems

N00014-06-C-0026

Submitted by

Lawrence Carin  
Signal Innovations Group  
1009 Slater Rd., Suite 200  
Durham, NC 27703

29 June 2006

REPORT DOCUMENTATION PAGE					Form Approved OMB No. 0704-0188	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p><b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b></p>						
1. REPORT DATE (DD-MM-YYYY) 06-29-2006		2. REPORT TYPE Progress Report		3. DATES COVERED (From - To) November 29, 2005 - June 29, 2006		
4. TITLE AND SUBTITLE Progress Report on Advanced Detection and Classification Algorithms for Acoustic-Color-Based Sonar Systems				5a. CONTRACT NUMBER N00014-06-C-0026		
				5b. GRANT NUMBER		
				5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) Dr. Lawrence Carin				5d. PROJECT NUMBER		
				5e. TASK NUMBER		
				5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Signal Innovations Group 1009 Slater Road, Suite 200 Durham, NC 27703				8. PERFORMING ORGANIZATION REPORT NUMBER CDRL A001		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Naval Research 875 N. Randolph Street, Suite 1425 Arlington, VA 22203-1995				10. SPONSOR/MONITOR'S ACRONYM(S) ONR		
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) CDRL A001		
12. DISTRIBUTION/AVAILABILITY STATEMENT Statement A						
13. SUPPLEMENTARY NOTES						
14. ABSTRACT Signal Innovations Group (SIG) has been working closely with the Naval Research Laboratory (NRL) on development of advanced algorithms for detection and classifying MCM targets, with data collected using the NRL sonar system. Over the current period of performance SIG has delivered to NRL kernel matching pursuits (KMP) software, that was employed by NRL at the most recent blind test. Details on the KMP algorithm are provided below. Additionally, NRL has recently delivered data from that blind test to SIG, and SIG is currently processing this data. NRL will soon be delivering to SIG data from their most recent sea test, for processing at SIG.						
15. SUBJECT TERMS						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON	
a. REPORT	b. ABSTRACT	c. THIS PAGE			Larry Carin	
U	U	U	UU	6	19b. TELEPHONE NUMBER (Include area code) (919) 323-3449	

## I. Progress Summary

Signal Innovations Group (SIG) has been working closely with the Naval Research Laboratory (NRL) on development of advanced algorithms for detection and classifying MCM targets, with data collected using the NRL sonar system. Over the current period of performance SIG has delivered to NRL kernel matching pursuits (KMP) software, that was employed by NRL at the most recent blind test. Details on the KMP algorithm are provided below. Additionally, NRL has recently delivered data from that blind test to SIG, and SIG is currently processing this data. NRL will soon be delivering to SIG data from their most recent sea test, for processing at SIG.

As detailed below, the KMP algorithm assume access to a set of separate training data, for the mines and clutter items of interest to the environment under test. This assumption was valid for the blind test the NRL executed. However, in many problems of practical importance, one may not have an appropriate set of training data, due to changes in the target and clutter characteristics, as well as changes to the channel properties. To address this problem SIG has been examining *in situ* learning algorithms, in which one integrates the sensing phase with classifier design. Details on this *in situ* learning algorithm (also termed active learning) are provided below.

## II. Kernel Matching Pursuits Details

We are interested in learning sparse kernel machines of functional form

$$f_n(\mathbf{x}) = \sum_{i=1}^n w_{n,i} K(\mathbf{c}_i, \mathbf{x}) + w_{n,0} = \mathbf{w}_n^T \boldsymbol{\phi}_n(\mathbf{x}) \quad (1)$$

where  $w_{n,0}$  is the bias term,  $K(\cdot, \cdot)$  is a kernel function measuring the similarity between two data samples

$$\boldsymbol{\phi}_n(\cdot) = [1, K(\mathbf{c}_1, \cdot), K(\mathbf{c}_2, \cdot), \dots, K(\mathbf{c}_n, \cdot)]^T \quad (2)$$

with  $\{y_i\}_{i=1}^N$  the kernel-induced basis function centered at  $\mathbf{c}_i$ , and

$$\mathbf{w}_n = [w_{n,0}, w_{n,1}, w_{n,2}, \dots, w_{n,n}]^T \quad (3)$$

are the weights that combine the basis functions in the summation, and the subscript  $n$  is used to denote the number of basis functions being used, with  $n < N$ . In the context of the binary classification problem consider in this section, a given  $\mathbf{x}$  is mapped to an estimated  $y \in \{0,1\}$  as  $y = U[f(\mathbf{x}) - 0.5]$ , where  $U(\alpha)$  is a unit step function, equal to one for  $\alpha \geq 0$ , and equal to zero otherwise. The form in (1) is the same as used in the SVM and RVM, although for the SVM  $K(\mathbf{c}_i, \mathbf{x})$  must be a Mercer kernel, while for the RVM and KMP this is not necessary.

The KMP implements a set of functions of the form in (1). Assume we are given a training set  $\{\mathbf{x}_i, y_i\}_{i=1}^N$ , where  $\mathbf{x}_i$  is the  $i^{\text{th}}$  input and  $y_i$  its expected output. The weighted sum of squared errors between the expected output and the KMP output given in (1) is

$$e_n = (1/\sum_{i=1}^N \beta_i) \sum_{i=1}^N \beta_i [y_i - f_n(\mathbf{x}_i)]^2 = (1/\sum_{i=1}^N \beta_i) \sum_{i=1}^N \beta_i [y_i - \mathbf{w}_n^T \boldsymbol{\phi}_n(\mathbf{x}_i)]^2 \quad (4)$$

where  $\beta_i$  is a constant responsible for quantifying the importance of the  $i$ th training sample  $(\mathbf{x}_i, y_i)$ . For example,  $1/\beta_i$  may represent the variance of the  $i$ th measurement; noisy measurements will therefore be given less importance when learning the model. In addition, if one has *a priori* knowledge that some data  $\mathbf{x}_i$  are in some sense “better” representative of the system being modeled this can be accounted for in the parameter  $\beta_i$ . The unknowns in (4) are the centers  $\mathbf{c}_i$  of the basis functions in  $\boldsymbol{\phi}_n$ , and the weights are represented by  $\mathbf{w}_n$ . The determination of  $\mathbf{c}_i$  is addressed separately below. At the moment we suppose  $\mathbf{c}_i$  and consequently  $\boldsymbol{\phi}_n$  are known and aim at solving for  $\mathbf{w}_n$ . Then the value of  $\mathbf{w}_n$  that minimizes (4) is found to be

$$\mathbf{w}_n = \mathbf{M}_n^{-1} \{\beta_i \boldsymbol{\phi}_{n,i} y_i\}_i \quad (5)$$

where  $\boldsymbol{\phi}_{n,i}$  is an abbreviation of  $\boldsymbol{\phi}_n(\mathbf{x}_i)$ ,  $\{\cdot\}_i \stackrel{\text{Def}}{=} \sum_{i=1}^N (\cdot)$ , and

$$\mathbf{M}_n = \sum_{i=1}^N \beta_i \boldsymbol{\phi}_n(\mathbf{x}_i) \boldsymbol{\phi}_n^T(\mathbf{x}_i) = \{\beta_i \boldsymbol{\phi}_{n,i} \boldsymbol{\phi}_{n,i}^T\}_i \quad (6)$$

is the Fisher information matrix. Note that for (6) to be a BLUE estimate, we have had to make no assumptions with regard to the statistics of  $y$  conditional on  $\mathbf{x}$ , other than that of a finite second moment.

An  $n$ th order KMP employs  $n$  basis functions. The  $(n+1)$ th order KMP is inductively written as

$$f_{n+1}(\mathbf{x}) = \mathbf{w}_{n+1}^T \boldsymbol{\phi}_{n+1}(\mathbf{x}) \quad (7)$$

where

$$\boldsymbol{\phi}_{n+1}(\cdot) = [1, K(\mathbf{c}_1, \cdot), K(\mathbf{c}_2, \cdot), \dots, K(\mathbf{c}_n, \cdot), K(\mathbf{c}_{n+1}, \cdot)]^T = \begin{bmatrix} \boldsymbol{\phi}_n(\cdot) \\ \boldsymbol{\phi}_{n+1}(\cdot) \end{bmatrix} \quad (8)$$

with  $\boldsymbol{\phi}_{n+1}(\cdot) = K(\mathbf{c}_{n+1}, \cdot)$  a new basis function centered at  $\mathbf{c}_{n+1}$ . The weighted sum of squared errors of the  $(n+1)$ th order KMP is

$$e_{n+1} = (1/\sum_{i=1}^N \beta_i) \sum_{i=1}^N \beta_i [y_i - f_{n+1}(\mathbf{x}_i)]^2 \quad (9)$$

Assuming the basis functions in  $\boldsymbol{\phi}_{n+1}$  are all known, then from (6)

$$\mathbf{w}_{n+1} = \mathbf{M}_{n+1}^{-1} \{\beta_i \boldsymbol{\phi}_{n+1,i} y_i\}_i \quad (10)$$

minimizes (14), where the Fisher information matrix  $\mathbf{M}_{n+1}$  is given as

$$\mathbf{M}_{n+1} = \{\beta_i^2 \boldsymbol{\phi}_{n+1,i} \boldsymbol{\phi}_{n+1,i}^T\}_i \quad (16)$$

One may show that  $\mathbf{w}_{n+1}$ , and  $e_{n+1}$  are respectively related to  $\mathbf{w}_n$  and  $e_n$  as

$$\mathbf{w}_{n+1} = \begin{bmatrix} \mathbf{w}_n + \mathbf{M}_n^{-1} \{\beta_i \boldsymbol{\phi}_{n,i} \boldsymbol{\phi}_{n+1,i}\}_i b^{-1} [\{\beta_i \boldsymbol{\phi}_{n,i}^T \boldsymbol{\phi}_{n+1,i}\}_i \mathbf{w}_n - \{\beta_i \boldsymbol{\phi}_{n+1,i} y_i\}_i] \\ -b^{-1} \{\beta_i \boldsymbol{\phi}_{n,i}^T \boldsymbol{\phi}_{n+1,i}\}_i \mathbf{w}_n + b^{-1} \{\beta_i \boldsymbol{\phi}_{n+1,i} y_i\}_i \end{bmatrix} \quad (11A)$$

$$e_{n+1} = e_n - \delta e(K, \mathbf{c}_{n+1}) \quad (11B)$$

where

$$\delta e(K, \mathbf{c}_{n+1}) = (1/\sum_{i=1}^N \beta_i) b^{-1} [\{\beta_i \phi_{n,i}^T \phi_{n+1,i}\}_i \mathbf{w}_n - \{\beta_i \phi_{n+1,i} y_i\}_i]^2 \quad (12)$$

and

$$b = \{\beta_i \phi_{n+1,i}^2\}_i - \{\beta_i \phi_{n+1,i} \phi_{n,i}^T\}_i \mathbf{M}_n^{-1} \{\beta_i \phi_{n,i} \phi_{n+1,i}\}_i \quad (13)$$

with  $\phi_{n+1,i} = K(\mathbf{c}_{n+1}, \mathbf{x}_i)$ .

Since  $\delta e(K, \mathbf{c}_{n+1})$  is dependent on the center  $\mathbf{c}_{n+1}$  of the new basis function, we obtain different values of  $\delta e(K, \mathbf{c}_{n+1})$  by selecting different  $\mathbf{c}_{n+1}$ . If we confine  $\mathbf{c}_{n+1}$  to be selected from the training data, we may conduct a “greedy” search in the training set but with the previously selected data excluded to avoid repetition. Formally, we have

$$\mathbf{c}_{n+1} = \mathbf{x}_{i_{n+1}} = \arg \max_{\substack{k \neq i_1, \dots, i_n \\ 1 \leq k \leq N}} \delta e(K, \mathbf{x}_k) \quad (14)$$

From (14)  $\delta e(K, \mathbf{c}_{n+1})$  depends on the functional form of the kernel  $K(\cdot, \cdot)$  as well as on support samples  $\mathbf{c}_{n+1}$ . This allows us to optimize the kernel to gain further error reduction. A simple approach to take is to first conduct a “greedy” search of  $\mathbf{c}_{n+1}$  in the training set, for a fixed kernel, and then fix  $\mathbf{c}_{n+1}$  and optimize the parameters of the kernel. For radial basis function (RBF) kernels, the only parameter other than  $\mathbf{c}_{n+1}$  is the kernel width, thus optimization of RBF kernels with  $\mathbf{c}_{n+1}$  fixed is a one-dimensional search for the kernel width. It is also possible to optimize  $\mathbf{c}_{n+1}$  and the kernel width simultaneously, but then  $\mathbf{c}_{n+1}$  is treated as a free parameter and is no longer confined to the training set. Another possibility is optimization over kernels of different functional forms, which offers greater diversity of the basis functions available to the KMP.

### III. In Situ Learning

Assume that the procedure discussed above selects  $n$  bases from the observed data  $\mathbf{X}$ . We now require labeled data to optimize the associated model weights  $\mathbf{w}$ . In a manner analogous to the previous discussion, we select those  $\mathbf{x}_i \in \mathbf{X}$  for which knowledge of the associated labels  $y_i$  would be most informative in the context of defining  $\mathbf{w}$ . Those  $\mathbf{x}_i$  that are so selected define a subset of signatures  $\mathbf{X}_s \subset \mathbf{X}$ , and these items are excavated to yield the respective set of labels  $\mathbf{L}_s$ . The set of signatures and labels  $(\mathbf{X}_s, \mathbf{L}_s)$  are then used to define the weights  $\mathbf{w}$  in a least-squares sense, and the resulting model  $f(\mathbf{x})$  is used to specify which of the remaining signatures  $\mathbf{x} \notin \mathbf{X}_s$  are likely targets of interest.

Assume that there are  $J$  signatures in  $\mathbf{X}_s$ , denoted  $\mathbf{X}_{s,J}$ . We quantify the information context in  $\mathbf{X}_{s,J}$  in the context of estimating the model weights  $\mathbf{w}$ , and further ask which  $\mathbf{x}_i \notin \mathbf{X}_{s,J}$  would be most informative if it and its label were added for determination of  $\mathbf{w}$ . We have

$$\mathbf{M}_n(\mathbf{X}_{s,J}) = \sum_{i: \mathbf{x}_i \in \mathbf{X}_{s,J}} \sigma_i^{-2} \phi_{n,i} \phi_{n,i}^T \quad (15)$$

The expressions (15) both employ an  $n$ -dimensional basis set  $\mathbf{B}_n \subset \mathbf{X}$ . In (15) the basis set  $\mathbf{B}_n$  is known and fixed, and we are only summing over those signatures  $\mathbf{X}_{s,J}$  for which knowledge of the associated labels is most informative in defining the model weights  $\mathbf{w}$ .

After adding a new signature  $\mathbf{x}_i \in \mathbf{X}$ ,  $\mathbf{x}_i \notin \mathbf{X}_{s,J}$ , we now have  $\mathbf{X}_{s,J+1}$  and  $\mathbf{M}_n$  is updated as

$$\mathbf{M}_n(\mathbf{X}_{s,J+1}) = \mathbf{M}_n(\mathbf{X}_{s,J}) + \sigma_{i_{J+1}}^{-2} \phi_{n,i_{J+1}} \phi_{n,i_{J+1}}^T \quad (16)$$

where  $i_{J+1}$  represents the index of the new signature selected for  $\mathbf{X}_{s,J+1}$ . Using the matrix identity  $\det(\mathbf{A} + \mathbf{F}\mathbf{F}^T) = \det(\mathbf{I} + \mathbf{F}^T\mathbf{A}^{-1}\mathbf{F})\det(\mathbf{A})$ , one obtains from (16)

$$q_n(\mathbf{X}_{s,J+1}) = q_n(\mathbf{X}_{s,J}) + \ln \rho(\mathbf{x}_{i_{J+1}}) \quad (17)$$

with

$$\rho(\mathbf{x}_{i_{J+1}}) = 1 + \sigma_{i_{J+1}}^{-2} \phi_{n,i_{J+1}}^T \mathbf{M}_n^{-1}(\mathbf{X}_{s,J}) \phi_{n,i_{J+1}} \quad (18)$$

Care is needed with regard to evaluating the inverse of  $\mathbf{M}_n$ , since if  $J < n$  the matrix is rank deficient. We have considered addressing this in either of two ways. A standard approach for inversion of such matrices is to add a small diagonal term to  $\mathbf{M}_n$ , such that its inverse exists. Alternatively, by construction one can assume that the items associated with the basis  $\mathbf{B}_n$  are all associated with  $\mathbf{X}_{s,J}$ , yielding a minimum of  $n$  labeled data and therefore assuring that the matrix is full rank. We have examined both procedures, and they yield comparable results.

Having addressed the inverse of  $\mathbf{M}_n$ , one iteratively maximizes  $\ln \rho(\mathbf{x}_{i_{J+1}})$  to obtain

$$\mathbf{x}_{i_{J+1}} = \arg \max_{\mathbf{x} \in \mathbf{X}, \mathbf{x} \notin \mathbf{X}_{s,J}} \ln \rho(\mathbf{x}) \quad (19)$$

Note that to define  $\mathbf{x}_{i_{J+1}}$  we again do not require the signature labels. The elements of  $\mathbf{X}_s$  are selected iteratively, in a “greedy” fashion as indicated in (19), until the information gain is below a prescribed threshold. After  $J$  iterations we have defined those signatures  $\mathbf{X}_{s,J}$  for which knowledge of the labels will best approximate the weights  $\mathbf{w}$ . These items are excavated, yielding the labels  $\mathbf{L}_{s,J}$ .

For the assumptions underlying the linear model in (1), and assuming knowledge of  $\mathbf{B}_n$  and  $(\mathbf{X}_{s,J}, \mathbf{L}_{s,J})$  the optimal estimation for the weights  $\mathbf{w}$  is expressed as

$$\mathbf{w} = [\Phi^T \Phi]^{-1} \Phi^T \mathbf{y} \quad (20)$$

where  $\mathbf{y}$  represents the set of labels determined via the  $J$  excavations

$$\mathbf{y} = \{y_{i_1}, y_{i_2}, \dots, y_{i_J}\}^T \quad (21)$$

and the  $J \times (n+1)$  matrix  $\Phi$  is defined as

$$\Phi = \begin{bmatrix} \phi_n^T(\mathbf{x}_{i_1}) \\ \phi_n^T(\mathbf{x}_{i_2}) \\ \vdots \\ \vdots \\ \vdots \\ \phi_n^T(\mathbf{x}_{i_J}) \end{bmatrix} \quad (22)$$

where, for example,  $\mathbf{x}_{i_1}$  corresponds to  $y_{i_1}$ .

In the classification stage we consider  $\mathbf{x} \notin \mathbf{X}_{s,J}$  and compute  $f(\mathbf{x})$ . For a prescribed threshold  $t$ ,  $\mathbf{x}$  is deemed associated with the +1 class if  $f(\mathbf{x}) \geq t$ , and associated with the -1 class if  $f(\mathbf{x}) < t$ , and by varying the threshold  $t$  one yields the receiver operating characteristic (ROC). The key component of the model  $f(\mathbf{x})$  is that it is linear in the weights  $\mathbf{w}$ , which yields a closed-form procedure for selection of  $\mathbf{B}_n$  and  $\mathbf{X}_{s,J}$ , as indicated in the previous sections.